Simplifying optimal strategies in limsup and liminf stochastic games

Discrete Applied Mathematics 251 (2018) 40–56

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Simplifying optimal strategies in limsup and liminf stochastic games☆

János Flesch [a], Arkadi Predtetchinski [b,*], William Sudderth [c]

[a] Department of Quantitative Economics, Maastricht University, P.O.Box 616, 6200 MD, The Netherlands
[b] Department of Economics, Maastricht University, P.O.Box 616, 6200 MD, The Netherlands
[c] School of Statistics, University of Minnesota, Minneapolis, MN 55455, United States

ARTICLE INFO

Article history:
Received 13 September 2016
Received in revised form 20 April 2018
Accepted 19 May 2018
Available online 19 June 2018

Keywords:
Zero-sum game
Stochastic game
Optimal strategy
Stationary strategy

ABSTRACT

We consider two-player zero-sum stochastic games with the limsup and with the liminf payoffs. For the limsup payoff, we prove that the existence of an optimal strategy implies the existence of a stationary optimal strategy. Our construction does not require the knowledge of an optimal strategy, only its existence. The main technique of the proof is to analyze the game with specific restricted action spaces. For the liminf payoff, we prove that the existence of a subgame-optimal strategy (i.e. a strategy that is optimal in every subgame) implies the existence of a subgame-optimal strategy under which the prescribed mixed actions only depend on the current state and on the state and the actions chosen at the previous period. In particular, such a strategy requires only finite memory. The proof relies on techniques that originate in gambling theory.

1. Introduction

Zero-sum stochastic games with the limsup and with the liminf payoffs. We consider two-player zero-sum stochastic games with finite state and action spaces. The game proceeds as follows: at the start of each period, the play is in one of the states. The two players then choose their actions simultaneously. The current state and the chosen actions determine an instantaneous reward that is to be paid by player 2 to player 1, and a probability distribution according to which the next state is drawn. Thus, the play of the game results in an infinite sequence of rewards, which is evaluated by a payoff function. Player 1's objective is to maximize the payoff, whereas player 2's objective is to minimize it. Zero-sum stochastic games have a wide range of applications not only in economics [1], but also in computer science (e.g. [2,5]), descriptive set theory and logic (e.g. [21]).

In this paper we examine the limsup and the liminf payoffs. The limsup payoff evaluates a sequence of rewards by the limit superior of the rewards, and so it evaluates by the peak performances. A player whose objective is to maximize the limsup payoff strives to obtain a good reward infinitely many times. The liminf payoff evaluates a sequence of rewards by the limit inferior of the rewards. A player whose objective is to maximize the liminf payoff strives to obtain bad rewards only finitely many times.

☆ The support for the visit of William Sudderth to Maastricht provided by a travel grant nr 040.11.495 of the Netherlands Organisation for Scientific Research (NWO) is acknowledged. The support of the COST Action CA16228 is acknowledged. We are grateful to two referees whose careful reading and useful suggestions have improved the exposition of our paper.

* Corresponding author.
E-mail addresses: j.flesch@maastrichtuniversity.nl (J. Flesch), a.predtetchinski@maastrichtuniversity.nl (A. Predtetchinski), bill@stat.umn.edu (W. Sudderth).

https://doi.org/10.1016/j.dam.2018.05.038
0166-218X/

Generally, in a zero-sum stochastic game with the limsup or with the liminf payoff, a player may have no optimal strategy at his disposal (cf. examples 13.4 and 13.5 in [19, p. 205]). The question that we address here is not whether an optimal strategy exists. Rather, assuming that an optimal strategy does exist, we ask how simple it can be. Since there is great interest, in theory as well as in applications, in strategies that only need finite memory, and especially in stationary strategies, we investigate whether the players have such optimal strategies.

Our main contributions. We prove the following simplification results.

I. For the limsup payoff, we prove that the existence of an optimal strategy implies the existence of a stationary optimal strategy. The proof relies on a construction of a specific type of stationary strategies, called maximally mixed. Roughly speaking, maximally mixed strategies are constructed as follows. To each state of the game we associate a "one-day" matrix game. In the one-day game the players choose their actions once, transition to a new state occurs, and the amount that player 1 receives from player 2 is equal to the value of the new state. Now consider mixed actions that are optimal in the one-day game. Among these mixed actions select those that have maximal support. A stationary strategy is called maximally mixed if, in every state, it prescribes such a mixed action.

The construction of maximally mixed strategies thus does not rely on the assumption that an optimal strategy exists. Rather, the assumption is only needed to prove that maximally mixed strategies are all optimal with respect to the limsup payoff.

II. For the liminf payoff, we prove the somewhat weaker statement that the existence of a subgame-optimal strategy (i.e. a strategy that is optimal in every subgame) implies the existence of a subgame-optimal strategy under which the prescribed mixed actions only depend on the current state and on the state and the actions chosen at the previous period. In particular, such a strategy requires only finite memory. The techniques used in the proof are very different, and mainly originate in gambling theory.

Related literature. The limsup and liminf payoffs have been analyzed in the game theoretic literature (e.g. [17–19,23], and [16]). These payoffs also play an important role in computer science. Chatterjee et al. [6] provide a survey of perfect information, or turn-based, stochastic games with the limsup and the liminf payoffs, both on the existential and the algorithmic aspects. Our model encompasses perfect information stochastic games as a special case: these are games where in every state only one of the players has more than one action.

Several payoff functions that are regularly considered in computer science can be written as a limsup or liminf payoff. For example, the payoff functions of reachability games and safety games can both be seen as limsup payoffs but also as liminf payoffs. The payoff functions of Büchi games can be seen as limsup payoffs, and those of co-Büchi games can be seen as liminf payoffs. For the definitions of these games, we refer to Chatterjee and Henziger [7].

We remark that results similar to ours are available in the literature of gambling and dynamic programming. Dubins and Savage [10] showed for a (one-person) gambling problem with a finite state space and limsup payoff that the existence of an optimal strategy implies the existence of a stationary optimal strategy. Blackwell [3] proved the same result for positive dynamic programming with a countable state space. There is also a generalization to a Borel measurable setting by Orkin [22]. Recently, Sudderth [26] showed for gambling problems with a countable state space and limsup payoff as well as for gambling problems with a finite state space and liminf payoff that the existence of an optimal strategy implies the existence of a Markov optimal strategy. A strategy is called Markov if the prescribed mixed actions only depend on the current state and on the current time period, but not directly on the past states and actions.

We now briefly discuss related results for payoffs other than the limsup and the liminf payoffs. For the discounted payoff, Shapley [24] showed that the players always have optimal strategies in the class of stationary strategies. For the average payoff, however, the players do not always have stationary optimal strategies, as was shown by the famous game called the Big Match (cf. [4,13], see also Example 1). Nevertheless, Flesch et al. [11] proved for the average payoff evaluation that if a player has an optimal strategy, then he also has a Markov optimal strategy.

The structure of the paper. The paper is organized as follows: Section 2 introduces the model, Section 3 is devoted to the description of limsup and liminf payoffs, Section 4 summarizes the main results, Sections 5 and 6 contain the proofs of the main results, and Section 7 discusses some extensions.

2. The model

Two-player zero-sum stochastic games. We examine two-player zero-sum stochastic games. Such a game is given by (1) a nonempty and finite state space S, (2) for each state $s \in S$, nonempty and finite action spaces $A(s)$ and $B(s)$ for player 1 and respectively player 2, (3) for each state $s \in S$ and actions $a \in A(s)$, $b \in B(s)$, a probability measure $p(s, a, b) = p(s'|s, a, b)_{s' \in S}$ on S, (4) a reward function $r : Z \to \mathbb{R}$, where $Z = \{(s, a, b)|s \in S, a \in A(s), b \in B(s)\}$, and (5) a payoff function $u : r(Z)^\infty \to \mathbb{R}$, where $r(Z)$ is the range of r.

We endow $r(Z)$ with the discrete topology, and $r(Z)^\infty$ with the product topology. We assume that u is bounded and Borel measurable.[1]

[1] The payoff function of the game could also be defined on the domain Z^∞. Note that the Borel measurability of u implies that the mapping $(z_0, z_1, \ldots) \to u(r(z_0), r(z_1), \ldots)$ from Z^∞ to $\mathbb{R}$ is also Borel measurable, when Z has the discrete topology and Z^∞ the product topology. For more details, we refer to Maitra and Sudderth [20].

The game is played at periods in $\mathbb{N} = \{0, 1, \ldots\}$ and begins in an initial state $s_0 \in S$. At every period $t \in \mathbb{N}$, the play is in a state $s_t \in S$. In this state, player 1 chooses an action $a_t \in A(s_t)$ and simultaneously player 2 chooses an action $b_t \in B(s_t)$. Then, with $z_t = (s_t, a_t, b_t)$, player 1 receives reward $r(z_t)$ from player 2, and state s_{t+1} is drawn in accordance with the probability measure $p(z_t)$. Thus, play of the game induces an infinite sequence $\{r(z_t)\}_{t\in\mathbb{N}}$ of rewards. The payoff is $u(r_0, r_1, \ldots,)$ where $r_t = r(z_t)$.

Strategies. The set of histories at period t is denoted by H_t. Thus, $H_0 = S$ and $H_t = Z^t \times S$ for every period $t \geq 1$. Let $H = \cup_{t\in\mathbb{N}} H_t$ denote the set of all histories. For each history h, let s_h denote the final state in h.

A mixed action for player 1 in state $s \in S$ is a probability measure $x(s)$ on $A(s)$. The support of $x(s)$, denoted by Support$(x(s))$, is defined as the set of actions in $A(s)$ having a positive probability under $x(s)$. Similarly, a mixed action for player 2 in state $s \in S$ is a probability measure $y(s)$ on $B(s)$. The support of $y(s)$, denoted by Support$(y(s))$, is defined as the set of actions in $B(s)$ having a positive probability under $y(s)$. The respective sets of mixed actions in state s are denoted by $X(s)$ and $Y(s)$.

A strategy for player 1 is a map π that to each history $h \in H$ assigns a mixed action $\pi(h) \in X(s_h)$. Similarly, a strategy for player 2 is a map σ that to each history $h \in H$ assigns a mixed action $\sigma(h) \in Y(s_h)$. The set of strategies is denoted by Π for player 1 and by Σ for player 2. A strategy is called pure if it places probability 1 on one action after each history.

A strategy is called stationary if the assigned mixed actions only depend on the history through its final state. Thus, a stationary strategy for player 1 can be seen as an element x of $X := \times_{s\in S} X(s)$. Similarly, a stationary strategy for player 2 can be seen as an element y of $Y := \times_{s\in S} Y(s)$. A pair of stationary strategies (x, y) induces a Markov chain on the state space S. A nonempty set $E \subseteq S$ is called ergodic with respect to (x, y), if starting in any state in E, the Markov chain eventually visits every state in E and visits no state outside E with probability 1.

A strategy is called Markov if the mixed action chosen in each period t depends only on the current state s_t together with t. Thus, a Markov strategy for player 1 can be seen as an element of $\times_{s\in S,\ t\in\mathbb{N}} X(s)$. Similarly, a Markov strategy for player 2 can be seen as an element $\times_{s\in S,\ t\in\mathbb{N}} Y(s)$. Markov strategies include stationary strategies as a special case.

For an initial state $s \in S$ and a pair of strategies $(\pi, \sigma) \in \Pi \times \Sigma$, we denote the expected payoff by $u(s, \pi, \sigma)$. Player 1's objective is to maximize the expected payoff given by u, and player 2's objective is to minimize it.

Value and optimality. It follows from the result of Martin [21] that the game has a value $v(s)$ for every initial state $s \in S$, i.e.

$$v(s) = \sup_{\pi\in\Pi} \inf_{\sigma\in\Sigma} u(s, \pi, \sigma) = \inf_{\sigma\in\Sigma} \sup_{\pi\in\Pi} u(s, \pi, \sigma).$$

For $\varepsilon \geq 0$, a strategy $\pi \in \Pi$ for player 1 is called ε-optimal for initial state s if $u(s, \pi, \sigma) \geq v(s) - \varepsilon$ for every strategy $\sigma \in \Sigma$ for player 2. Similarly, a strategy $\sigma \in \Sigma$ for player 2 is called ε-optimal for initial state s if $u(s, \pi, \sigma) \leq v(s) + \varepsilon$ for every strategy $\pi \in \Pi$ for player 1. A strategy is called ε-optimal if it is ε-optimal for every initial state.

In view of the existence of the value, for all $\varepsilon > 0$, each player has an ε-optimal strategy. A 0-optimal strategy is simply called optimal.

Subgame-optimality. Consider a strategy $\pi \in \Pi$ for player 1 and a sequence $g \in Z^t$, for some $t \in \mathbb{N}$ (for $t = 0$, g is the empty sequence). The continuation strategy $\pi|_g$ is the strategy in Π given, for every history $h \in H$, by $\pi|_g(h) = \pi(gh)$, where gh stands for the concatenation of g and h.

A strategy $\pi \in \Pi$ for player 1 is called subgame-optimal, if $\pi|_g$ is optimal for every $g \in Z^t$ and $t \in \mathbb{N}$. Continuation strategies and subgame-optimal strategies for player 2 are defined analogously.

Note that every subgame-optimal strategy is optimal. The converse holds for stationary strategies: a stationary optimal strategy is always subgame-optimal.

3. The limsup and liminf payoffs, and examples

In this section we define two closely related payoff functions, the limsup and the liminf payoffs, and provide some examples.

3.1. The limsup payoff

The limsup payoff is defined for each $(r_0, r_1, \ldots) \in r(Z)^\infty$ as

$$\overline{u}(r_0, r_1, \ldots) = \limsup_{t\to\infty} r_t.$$

For every $s \in S$, let $\overline{v}(s)$ denote the limsup value for initial state s.

We emphasize that, under our definition of the evaluation $\overline{u}(s, \pi, \sigma)$ of the pair of strategies (π, σ) in the initial state s, one computes the expectation of the limit superior of the sequence of rewards, rather than the limit superior of the expectation of the rewards. We briefly discuss the latter evaluation in Section 7.

For the limsup payoff, player 1 generally does not have a stationary ε-optimal strategy, for small $\varepsilon > 0$. This is shown by example 13.4 in [19, p. 205]. We illustrate this point by means of an adaptation of the well-known Big Match.

Example 1. Consider the following game[2] with the limsup payoff $\overline{u}$. There are three states $S = \{s, 0^*, 1^*\}$. State 0^* is absorbing, and the reward is 0 in this state. State 1^* is also absorbing, and the reward is 1 in this state. State s is the interesting state. In this state, the actions are T and B for player 1, and L and R for player 2. Action pair (T, L) gives reward 1 and the state remains s, action pair (T, R) gives reward 0 and the state remains s, action pair (B, L) gives reward 0 and leads to state 0^*, and action pair (B, R) gives reward 1 and leads to state 1^*. The game can be represented as follows, where $*$ denotes absorption.

	L	R
T	1	0
B	0^*	1^*

We prove the following statements for initial state s:

(a) The value is $\overline{v}(s) = 1$.
(b) Player 1 has no optimal strategy.
(c) Player 1 has no stationary ε-optimal strategy for $\varepsilon \in [0, 1)$.
(d) Player 1 has no Markov ε-optimal strategy for $\varepsilon \in [0, \frac{1}{2})$.

We use the notations L^∞ and R^∞ for the strategies of player 2 that always choose action L and action R in state s, respectively.

Proof of (a): Take an $\varepsilon \in (0, 1)$, and a sequence $(w_n)_{n\in\mathbb{N}}$ in $(0, 1)$ such that $\Pi_{n=0}^{\infty} w_n = 1 - \varepsilon$. Let π^ε be the strategy for player 1 that at period t prescribes action T with probability $w_{\ell(t)}$ and action B with probability $1 - w_{\ell(t)}$, where $\ell(t)$ denotes the number of times action L has been played by player 2 before period t. We prove that for every strategy τ for player 2

$$\overline{u}(s, \pi^\varepsilon, \tau) \geq 1 - \varepsilon. \tag{1}$$

Indeed, take any pure strategy τ for player 2. (Given player 1's strategy π^ε, player 2 is facing a Markov Decision Problem, so it is sufficient to consider only pure responses for player 2). Such a strategy gives rise to a sequence $\tau = (b_t)_{t\in\mathbb{N}}$, where $b_t \in \{L, R\}$ is the action player 2 takes in period t, given that player 1 has not played action B thus far. Consider the play under (π^ε, τ). Note that the total probability of absorption in entry (B, L) is at most ε under (π^ε, τ). Indeed, the total probability of absorption in entry (B, L) under (π^ε, τ) is at most the total probability of absorption in entry (B, L) when player 2 always chooses action L against π^ε, and this latter probability is exactly $\varepsilon = 1 - \Pi_{n=0}^{\infty} w_n$ by the choice of π^ε. We distinguish two cases.

Suppose first that L occurs infinitely often in τ. Since the total probability of absorption in entry (B, L) is at most ε, with probability of at least $1 - \varepsilon$ one of the following events occurs: either absorption takes place at some period in entry (B, R), or entry (T, L) is played infinitely often. Hence, (1) follows.

Now assume that L occurs only finitely many times in τ. Since the total probability of absorption in entry (B, L) is at most ε, the probability that absorption in entry (B, R) eventually occurs is at least $1 - \varepsilon$. Hence, (1) follows again.

Thus, $\overline{v}(s) = 1$ indeed, and the strategy π^ε is ε-optimal for player 1, for every $\varepsilon \in (0, 1)$.

Proof of (b)[3]: Take any strategy π for player 1. We prove that π is not optimal by distinguishing two cases.

Case 1: the probability under (s, π, R^∞) that action B is ever chosen is zero. In this case, $\overline{u}(s, \pi, R^\infty) = 0$, and π is not optimal.

Case 2: the probability under (s, π, R^∞) that action B is ever chosen is positive. Let t denote the first period at which action B is chosen with a positive probability under (s, π, R^∞). Consider any strategy σ for player 2 which chooses action R before period t and chooses action L at period t. Then, $\overline{u}(s, \pi, \sigma) < 1$, and π is not optimal in this case either.

Proof of (c): Take any stationary strategy x for player 1. If x places probability 1 on action T, then $\overline{u}(s, x, R^\infty) = 0$. Otherwise, if x places probability less than 1 on action T, then $\overline{u}(s, x, L^\infty) = 0$. Thus, x is not ε-optimal for $\varepsilon \in [0, 1)$, as claimed.

Proof of (d): Let π be a Markov strategy for player 1 and, for $t \in \mathbb{N}$, let w_t be the probability that π assigns to action T when in state s in period t. Fix $\varepsilon \in [0, \frac{1}{2})$.

Case 1: $\Pi_{t=0}^{\infty} w_t > \varepsilon$. The probability under (s, π, R^∞) that the state remains equal to s and the players choose actions (T, R) forever is greater than ε. Hence, $\overline{u}(s, \pi, R^\infty) < 1 - \varepsilon$.

Case 2: $\Pi_{t=0}^{\infty} w_t \leq \varepsilon$. The probability under (s, π, L^∞) that player 1 eventually plays B is at least $1 - \varepsilon$. Hence, $\overline{u}(s, \pi, L^\infty) \leq \varepsilon$. Since $\varepsilon < \frac{1}{2}$, we have $\varepsilon < 1 - \varepsilon$, and the proof is complete. ◇

Remark. In the above game, the ε-optimal strategy for player 1 constructed in part (a) takes into account which actions player 2 has chosen in the past. However, if the rewards only depend on the current state but not on the actions, then player 1 always has ε-optimal strategies that only take the sequence of states visited into account when recommending a mixed action. This can be verified by following the steps in the iterative procedure for calculating the value of the game given in [17–19]. This procedure is, in general, transfinite for games with an infinite state space. However, it is simpler for games with a finite state space like those here (Theorem 7.11.13, page 201 in [19]).

[2] This game, when the average payoff is considered instead of the limsup payoff, is the famous Big Match, which was introduced by Gillette [13]. The average payoff evaluates a sequence of rewards by taking the long-term average reward. It was shown by Blackwell and Ferguson [4] that the game has a value for the average payoff. We also refer to section 7.17 in [19].

[3] A similar proof is credited to Dubins by Blackwell and Ferguson [4] for the Big Match with the average payoff.

3.2. The liminf payoff

The liminf payoff is defined for each $(r_0, r_1, \dots) \in r(Z)^\infty$ as

$$\underline{u}(r_0, r_1, \dots) = \liminf_{t\to\infty} r_t.$$

For every $s \in S$, let $\underline{v}(s)$ denote the liminf value for initial state s.

We emphasize that, under our definition of the evaluation $\underline{u}(s, \pi, \sigma)$ of the pair of strategies (π, σ) in the initial state s, one computes the expectation of the limit inferior of the sequence of rewards, rather than the limit inferior of the expectation of the rewards. We briefly discuss the latter evaluation in Section 7.

The limsup and the liminf payoffs are closely related. Indeed, in a zero-sum stochastic game with the limsup payoff, player 1's objective is to maximize the limsup payoff, and player 2's objective is to minimize it. The objective of player 2 can however also be seen as maximizing the liminf of the rewards given by $-r$, the negative of the reward function.[4] Similarly, in a zero-sum stochastic game with the liminf payoff, the objective of player 2 can also be seen as maximizing the limsup of the rewards given by $-r$.

Example 2. We illustrate that the limsup and the liminf values can be different. For this purpose, we revisit Example 1, but now with the liminf payoff.

	L	R
T	1	0
B	0*	1*

For any $\varepsilon \in (0, 1)$, player 2 can guarantee that the liminf payoff is at most ε, by using the stationary strategy that chooses action L with probability $1 - \varepsilon$ and action R with probability ε. Hence, the liminf value for state s is $\underline{v}(s) = 0$. Recall that the limsup value for state s is $\overline{v}(s) = 1$. ◇

Secchi [23] showed for the liminf payoff that player 1 always has a stationary ε-optimal strategy, for every $\varepsilon > 0$. The following game, which is a variation of example 13.5 in [19, p. 205], shows that player 1 generally does not have a 0-optimal strategy for the liminf payoff.

Example 3. The game and the notation are similar to those in Example 1. The only modification is that entry (T, L) in state s is also absorbing.

	L	R
T	1*	0
B	0*	1*

We consider the liminf payoff. For any $\varepsilon \in (0, 1)$, player 1 can guarantee a liminf payoff of at least $1 - \varepsilon$, by using the stationary strategy x^ε that chooses action T with probability $1 - \varepsilon$ and action B with probability ε. Indeed, the strategy x^ε makes sure that absorption eventually takes place and that the total probability of absorption in entry (B, L) is at most ε, regardless of the strategy of player 2. Hence, the liminf value is $\underline{v}(s) = 1$. However, player 1 has no optimal strategy, which can be verified similarly to claim (b) of Example 1. ◇

4. The main results

In this section, we discuss our main results. We start with the limsup payoff.

Theorem 1 (*Limsup Payoff*)**.** *Consider a zero-sum stochastic game with the limsup payoff. If player 1 has an optimal strategy, then he has a stationary optimal strategy as well.*

The proof of the theorem borrows a number of ideas from Flesch et al. [11]. Associated to each state of the game is a "one-day" matrix game. In the one-day game the players choose their actions once, the transition to a new state occurs, and the amount that player 1 receives from player 2 is equal to the value of the new state. We then consider mixed actions for player 1 that are optimal in the one-day game. Of these mixed actions we select those that have the largest support. A stationary strategy for player 1 is called maximally mixed if, in every state, it prescribes such a mixed action.

This construction does not rely on the assumption that an optimal strategy exists for player 1. Rather, the assumption is invoked at a later stage of the proof, when we show that all maximally mixed strategies for player 1 are optimal with respect to the limsup payoff.

Recall that every stationary optimal strategy is subgame-optimal. Hence, by the above theorem, if player 1 has an optimal strategy in a limsup game then he also has a subgame-optimal strategy.

For the liminf evaluation, we obtain the following simplification result.

[4] Here, we use that $\limsup_{t\to\infty} r(z_t) = -\liminf_{t\to\infty} -r(z_t)$, so minimizing $\limsup_{t\to\infty} r(z_t)$ is equivalent to maximizing $\liminf_{t\to\infty} -r(z_t)$.

Theorem 2 (*Liminf Payoff*)**.** *Consider a zero-sum stochastic game with the liminf payoff. If player 1 has a subgame-optimal strategy, then he also has a subgame-optimal strategy under which the prescribed mixed actions only depend on the current state and on the state and the actions chosen at the previous period.*

The proof relies on techniques that originate in gambling theory (e.g. [23,25]. It remains an open problem whether the assumption in Theorem 2 can be weakened to the existence of an optimal strategy, and whether one can then derive the existence of a stationary optimal strategy.

In Section 7, we discuss extensions of the main results.

5. The proof of Theorem 1

In this section we prove Theorem 1. First we define maximally mixed strategies for player 1. Then, we define the auxiliary concept of effective strategies for player 1. Subsequently, we show that effective strategies are optimal, given that player 1 has an optimal strategy. Finally, we show that maximally mixed strategies are effective. The final subsection provides several illustrative examples.

5.1. Maximally mixed strategies

Consider a zero-sum stochastic game G with the limsup payoff. For each state $s \in S$, we consider a matrix game $M(s)$. This matrix game is fundamental and frequently used in stochastic games.

Let $s \in S$. The matrix game $M(s)$ is defined as follows. The sets of actions are $A(s)$ and respectively $B(s)$ for players 1 and 2, and the payoff for each pair of actions $(a, b) \in A(s) \times B(s)$ is

$$u_M(s, a, b) := \sum_{s' \in S} p(s'|s, a, b) \cdot \overline{v}(s').$$

Note that $u_M(s, a, b)$ is the expectation of the limsup value after transition in the original game G, when the players play actions a and b in state s.

The value of the matrix game $M(s)$ is exactly $\overline{v}(s)$:

$$\mathrm{Val}(M(s)) = \overline{v}(s).$$

Indeed, if π denotes an ε-optimal strategy for player 1 in the game G for the initial state s, then $u_M(s, \pi(s), y) \geq \overline{v}(s) - \varepsilon$ must hold for every $y \in Y(s)$, where $\pi(s)$ is the mixed action that π prescribes for the initial state s. This implies that $\mathrm{Val}(M(s)) \geq \overline{v}(s)$. We similarly obtain the opposite inequality.

In the matrix game $M(s)$, we define

$$X^*(s) := \{x \in X(s) \mid u_M(s, x, y) \geq \overline{v}(s) \ \ \forall y \in Y(s)\}$$

$$Y^*(s) := \{y \in Y(s) \mid u_M(s, x, y) = \overline{v}(s) \ \ \forall x \in X^*(s)\}.$$

The set $X^*(s)$ consists of all optimal mixed actions for player 1 in the matrix game $M(s)$, whereas $Y^*(s)$ is the set of equalizers for player 2. Both $X^*(s)$ and $Y^*(s)$ are nonempty polytopes. In fact, all optimal mixed actions of player 2 in the matrix game $M(s)$ belong to $Y^*(s)$.

Define

$$A^*(s) := \{a \in A(s) \mid x(a) > 0 \text{ for some } x \in X^*(s)\},$$

and

$$B^*(s) := \{b \in B(s) \mid y(b) > 0 \text{ for some } y \in Y^*(s)\}.$$

The set $A^*(s)$, respectively $B^*(s)$, consists of all actions in $A(s)$, respectively $B(s)$, that are used by some mixed action in $X^*(s)$, respectively $Y^*(s)$.

It is easy to see that if $y \in Y^*(s)$ then $\mathrm{Support}(y) \subseteq Y^*(s)$. Indeed, let $x \in X^*(s)$. Then

$$\overline{v}(s) = u_M(s, x, y) = \sum_{b \in B} y(b) u_M(s, x, b).$$

We have $u_M(s, x, b) \geq \overline{v}(s)$ for each $b \in B$. Therefore, if $y(b) > 0$ then $u_M(s, x, b) = \overline{v}(s)$, as desired. It now follows that $B^*(s) \subseteq Y^*(s)$. Since $Y^*(s)$ is convex, we also have

$$Y^*(s) = \{y \in Y(s) \mid \mathrm{Support}(y) \subseteq B^*(s)\}.$$

Define

$$X^{**}(s) := \{x \in X^*(s) \mid x(a) > 0 \ \ \forall a \in A^*(s)\}.$$

The set $X^{**}(s)$ consists of all mixed actions in $X^*(s)$ which put positive probability on each such action. By convexity of $X^*(s)$, the set $X^{**}(s)$ is nonempty.[5]

Define $X^{**} = \times_{s\in S} X^{**}(s)$. Thus, the set X^{**} consists of all stationary strategies for player 1 that use a mixed action in $X^{**}(s)$ in every state $s \in S$. These stationary strategies are called *maximally mixed* strategies.

The following lemma is closely related to lemma 2.5 in [11]. The first part of the lemma is a well known fact that for a pair of stationary strategies the limsup payoff is constant on each ergodic set. The second and the third parts claim that if player 1 uses a maximally mixed strategy and player 2 uses a stationary strategy, then on each ergodic set, the limsup value is a constant and player 2's strategy behaves as a strategy that always chooses mixed actions in $Y^*(s)$ in each state s. The intuition behind this result is as follows. Observe that starting from any state in an ergodic set E, all the states in E must be visited infinitely often and hence all the triples (s, a, b) with positive probability must also occur infinitely often with probability one. Thus, the limsup payoff must almost surely equal the maximum of $r(s, a, b)$ taken over all such triples and so $\overline{v}(s')$ must also equal this maximum for all s' in E. In particular, the limsup value is constant on E.

For $(x, y) \in X \times Y$ and $E \subseteq S$ define

$$\overline{u}(E, x, y) = \max\{r(s, a, b) \mid s \in E,\ x(s)(a) > 0,\ y(s)(b) > 0\}.$$

Lemma 3. *Let $E \subseteq S$ be an ergodic set for some pair of stationary strategies $(x, y) \in X \times Y$.*

1. *$\overline{u}(s, x, y) = \overline{u}(E, x, y)$ for each $s \in E$.*
2. *Suppose that $x \in X^{**}$. Then $\overline{v}(s) = \overline{v}(s')$ for every $s, s' \in E$. Henceforth we denote this constant by $\overline{v}(E)$.*
3. *Suppose that $x \in X^{**}$. Then $y(s) \in Y^*(s)$ for every $s \in E$.*

Proof. Consider the Markov chain on the states induced by (x, y).

[1] Starting from any state s in the ergodic set E, all the states in E must be visited infinitely often and hence all the triples (s', a, b) with positive probability must also occur infinitely often with probability one. Hence the payoff $\overline{u}(s, x, y)$ equals the maximal reward over all such triples.

[2] Let $x \in X^{**}$. Recall that $x(s) \in X^{**}(s) \subseteq X^*(s)$ for every $s \in E$. Let $\overline{v}(E) = \max_{s\in E} \overline{v}(s)$. Consider the set $E' = \{s \in E \mid \overline{v}(s) = \overline{v}(E)\}$. We argue that $E' = E$. Suppose $s \in E'$. It follows from the inequality $u_M(s, x, y) \geq \overline{v}(s)$ and the equality

$$u_M(s, x, y) = \sum_{s',a,b} p(s'|s, a, b) \cdot \overline{v}(s') \cdot x(s)(a) \cdot y(s)(b)$$

that the next state s' is in E' with probability 1. Thus the Markov chain never leaves the set E'. However, every state in the ergodic set E must be visited almost surely starting from any state in $E' \subseteq E$. We conclude that $E' = E$.

[3] Let $x \in X^{**}$. Take a state $s \in E$ and a mixed action $x'(s) \in X^*(s)$. We know that $(x(s), y(s))$ only induces transitions to states in E. Since $x(s) \in X^{**}(s)$, we have that the support of $x(s)$ includes the support of $x'(s)$. Hence $(x'(s), y(s))$ also induces transitions to states in E only. Therefore by part [2], $u_M(s, x'(s), y(s)) = \overline{v}(E) = \overline{v}(s)$. Since $x'(s)$ is chosen arbitrarily in $X^*(s)$, we have shown that $y(s) \in Y^*(s)$, as desired. □

In the remaining part of this section, we prove the following result, which in particular implies Theorem 1.

Theorem 4. *Consider a zero-sum stochastic game with the limsup payoff. Assume that player 1 has an optimal strategy. Then every maximally mixed strategy is optimal for player 1.*

Note that the construction of the set X^{**} of maximally mixed strategies does not rely on the existence of an optimal strategy. As we will see in the subsequent subsections, the assumption that there exists an optimal strategy for player 1 will only be invoked to guarantee that the strategies in X^{**} are optimal. We emphasize that we do not need to know an optimal strategy, its existence is sufficient.

We remark that the set X^{**} also played an important role in [11]. They proved with respect to the average payoff that, if player 1 has an optimal strategy, then for every $\varepsilon > 0$, he has a stationary ε-optimal strategy within X^{**}, and that he also has a Markov optimal strategy that only makes use of mixed actions in $X^{**}(s)$, for all $s \in S$.

We prove Theorem 4 in two main steps. First, we introduce one other class of stationary strategies, called effective, and show that if player 1 has an optimal strategy then all effective strategies are optimal (Theorem 5). Then, we prove that all maximally mixed strategies are effective (Theorem 6).

5.2. Effective strategies

Restricted sets of strategies. Let Π^* denote the set of strategies π for player 1 such that $\pi(h) \in X^*(s_h)$ for each history $h \in H$. Let $X^* = \times_{s\in S} X^*(s)$ denote the set of stationary strategies in Π^*. Note that $X^{**} \subseteq X^*$.

[5] Any convex combination of all extreme points of the polytope $X^*(s)$ with only positive weights belongs to $X^{**}(s)$.

Similarly, let Σ^* denote the set of strategies σ for player 2 such that $\sigma(h) \in Y^*(s_h)$ for each history $h \in H$, and let $Y^* = \times_{s\in S} Y^*(s)$ denote the set of stationary strategies in Σ^*.

Effective strategies. For every state $s \in S$, we define

$$\gamma^*(s) := \sup_{\pi\in\Pi^*} \inf_{\sigma\in\Sigma^*} \overline{u}(s,\pi,\sigma).$$

Intuitively, $\gamma^*(s)$ is the highest expected payoff that player 1 can guarantee from state s by using a strategy in Π^*, given that player 2 is restricted to using a strategy in Σ^*. As we show below (see the proof of Theorem 5), if player 1 has an optimal strategy and he is using it against a strategy of player 2 from Σ^*, then after any history h that occurs with a positive probability, the optimal strategy of player 1 recommends a mixed action in $X^*(s_h)$. This means intuitively that, against any strategy in Σ^*, the optimal strategy of player 1 behaves as a strategy in Π^*. This will also imply that, if player 1 has an optimal strategy, then $\gamma^*(s)$ is at least as large as the limsup value of the original game $\overline{v}(s)$.

A maximally mixed strategy $x \in X^{**}$ for player 1 is called *effective*, if player 2 has a stationary strategy $y \in Y$ with the following two properties:

1. The strategy y is a best response for the limsup payoff against x, i.e. $\overline{u}(s,x,y) \le \overline{u}(s,x,\sigma)$ for every state $s \in S$ and every strategy σ for player 2.
2. For every ergodic set $E \subseteq S$ with respect to (x,y) we have $\overline{u}(E,x,y) \ge \min_{s\in E}\gamma^*(s)$.

It will follow from Theorem 6 that, in every zero-sum stochastic game with the limsup payoff, player 1 has an effective strategy.

5.3. Optimality of effective strategies

The following result is shown in [11] for the case of the average payoff. The main ideas of the proof remain the same for the limsup payoff.

Theorem 5. *Consider a zero-sum stochastic game with the limsup payoff. Assume that player 1 has an optimal strategy. Then, every effective strategy for player 1 is optimal.*

Proof. Step 1 (cf. lemma 2.2 in [11]): Let $\pi \in \Pi$ be an optimal strategy for player 1, $\sigma \in \Sigma^*$ be a strategy for player 2, and $s \in S$ be the initial state. Then, for every history $h \in H$ with $\mathbb{P}_{(s,\pi,\sigma)}(h) > 0$, we have $\pi(h) \in X^*(s_h)$. This means intuitively that, against any strategy in Σ^*, the strategy π behaves as a strategy in Π^*.
Proof of step 1: Indeed, suppose by way of contradiction that there is a history $h \in H_t$, for some $t \in \mathbb{N}$, for which we have $\mathbb{P}_{(s,\pi,\sigma)}(h) > 0$ and $\pi(h) \notin X^*(s_h)$. Assume that h is a shortest such history. That means that, with probability 1 with respect to (s,π,σ), at each period $t' < t$, π prescribes a mixed action in $X^*(s_{t'})$ and σ prescribes a mixed action in $Y^*(s_{t'})$. Consequently, the sequence $\overline{v}(s) = \overline{v}(s_0), \overline{v}(s_1), \ldots, \overline{v}(s_t)$ is a martingale (up to period t), with respect to $\mathbb{P}_{(s,\pi,\sigma)}$, so that in particular

$$\mathbb{E}_{(s,\pi,\sigma)}(\overline{v}(s_t)) = \overline{v}(s). \tag{2}$$

For $\delta > 0$, let σ^δ be a strategy for player 2 such that: (i) before period t: σ^δ agrees with σ, (ii) at period t: σ^δ plays a best response to $\pi(h_t)$ in the matrix game $M(s_t)$, and (iii) from $t+1$ on: σ^δ plays a δ-optimal strategy from the state s_{t+1}. Due to property (i) and equality (2), for all $\delta > 0$

$$\mathbb{E}_{(s,\pi,\sigma^\delta)}(\overline{v}(s_t)) = \mathbb{E}_{(s,\pi,\sigma)}(\overline{v}(s_t)) = \overline{v}(s). \tag{3}$$

Due to property (ii) and the mistake that π makes at h, it follows for all $\delta > 0$ sufficiently small that

$$\mathbb{E}_{(s,\pi,\sigma^\delta)}(\overline{v}(s_{t+1})) + \delta < \mathbb{E}_{(s,\pi,\sigma^\delta)}(\overline{v}(s_t)). \tag{4}$$

Due to property (iii), for all $\delta > 0$

$$\overline{u}(s,\pi,\sigma^\delta) \le \sum_{s'} \mathbb{P}_{(s,\pi,\sigma^\delta)}(s_{t+1} = s') \cdot (\overline{v}(s') + \delta) = \mathbb{E}_{(s,\pi,\sigma^\delta)}(\overline{v}(s_{t+1})) + \delta. \tag{5}$$

Combining (5), (4), and (3), we obtain that for all $\delta > 0$ sufficiently small,

$$\overline{u}(s,\pi,\sigma^\delta) < \overline{v}(s).$$

This however contradicts the optimality of π.

Step 2 (cf. lemma 2.3 in [11]): $\gamma^*(s) \ge \overline{v}(s)$ for every state $s \in S$. This means intuitively that it weakly favors player 1 if the players are restricted to the strategy sets Π^* and Σ^*.
Proof of step 2: This step is a direct consequence of the assumption of Theorem 5 and step 1 above, as any optimal strategy of player 1 behaves as a strategy in Π^* against a strategy in Σ^*.

Step 3: Conclusion of the proof of Theorem 5. Let x be an effective strategy for player 1. Then there exists a best response $y \in Y$ to x such that for every ergodic set E with respect to (x, y) we have $\overline{u}(E, x, y) \geq \min_{s\in E}\gamma^*(s)$. We now argue that $\overline{u}(w, x, y) \geq \overline{v}(w)$ for every initial state $w \in S$ and thus establish the optimality of x.

First consider the case when the initial state w belongs to some ergodic set E for (x, y). By part 2 of Lemma 3, the limsup value on E is a constant $\overline{v}(E)$. Thus, we need to show that $\overline{u}(E, x, y) \geq \overline{v}(E)$. By the definition of an effective strategy together with step 2, we have

$$\overline{u}(E, x, y) \geq \min_{s\in E} \gamma^*(s) \geq \min_{s\in E} \overline{v}(s) = \overline{v}(E),$$

as desired.

Now suppose that the initial state w is transient for the Markov chain on S determined by (x, y). Thus w does not belong to an ergodic set. However, the random stopping time τ at which the process $s_0 = w, s_1, s_2, \ldots$ first reaches an ergodic set is finite with $\mathbb{P}_{(w,x,y)}$-probability one. Also because $x(s) \in X^{**}(s) \subseteq X^*(s)$ for every $s \in S$, the process $\overline{v}(s_0), \overline{v}(s_1), \ldots$ is a bounded submartingale (and even a martingale, as y is a best response to x). Hence,

$$\overline{v}(w) \leq \mathbb{E}_{(w,x,y)}\overline{v}(s_\tau) \leq \mathbb{E}_{(w,x,y)}\overline{u}(s_\tau, x, y) = \overline{u}(w, x, y).$$

The first inequality above is by the optional sampling theorem for bounded submartingales; the second is by the previous case since s_τ belongs to an ergodic set $\mathbb{P}_{(w,x,y)}$-almost surely; the equality follows by conditioning on the history up to time τ and using the fact that the limsup of the sequence $r_0, r_1, \ldots$ is the same as the limsup of the sequence $r_\tau, r_{\tau+1}, \ldots$ $\mathbb{P}_{(w,x,y)}$-almost surely.

This completes the proof of Theorem 5. □

5.4. Maximally mixed strategies are effective

Theorem 6. *Consider a zero-sum stochastic game with the limsup payoff. Then, every maximally mixed strategy is effective.*

Proof. Take any $x \in X^{**}$. Once x has been fixed, player 2 essentially faces a Markov Decision Problem. Consequently, player 2 has a stationary best response $y \in Y$ to x, in fact even a pure one (cf. [25], as well as [6], where an argument is given based on [8,14] and [9]; see also [15]).

Let E be an ergodic set for (x, y). We show that

$$\overline{u}(E, x, y) \geq \gamma^*(s) \quad \forall s \in E. \tag{6}$$

By part (3) of Lemma 3, $y(s) \in Y^*(s)$ for all $s \in E$. Choose any $y' \in Y^*$ that coincides with y on E. Clearly we have

$$\overline{u}(E, x, y') = \overline{u}(E, x, y). \tag{7}$$

Now for each $\pi \in \Pi^*$ and each $s \in E$ it holds that $\overline{u}(E, x, y') \geq \overline{u}(s, \pi, y')$. To see this, take any $\pi \in \Pi^*$ and $s \in E$. Since $x \in X^{**}$, if π places positive probability on an action $a \in A(s_h)$ after a history h, then this action a is also played with a positive probability under x. Consequently, if, starting with the state s, the state s' is reached with positive probability under (π, y'), it is reached with positive probability under (x, y'), and hence also under (x, y). It follows that s' is an element of E. Now the claim follows in view of Eq. (7).

Thus we conclude that for each $s \in E$,

$$\overline{u}(E, x, y) = \overline{u}(E, x, y') \geq \sup_{\pi\in\Pi^*} \overline{u}(s, \pi, y') \geq \sup_{\pi\in\Pi^*} \inf_{\sigma\in\Sigma^*} \overline{u}(s, \pi, \sigma) = \gamma^*(s),$$

which proves (6). □

5.5. Examples

Example 4 (*Repeated Games*)**.** When the set of states is a singleton, i.e. when $S = \{s\}$, we essentially have a repeated game. In this case the associated matrix game is trivial in the sense that $u_M(s, a, b)$ is equal to the value $\overline{v}(s)$ for each $(a, b) \in A(s)\times B(s)$. Consequently, $X^*(s) = X(s)$, $X^{**}(s) = \{x \in X(s) : x(a) > 0 \text{ for each } a \in A(s)\}$ and $Y^*(s) = Y(s)$. Further, $\Pi^* = \Pi$ and $\Sigma^* = \Sigma$.

Let

$$R = \min_{b\in B(s)} \max_{a\in A(s)} r(s, a, b). \tag{8}$$

The limsup value of the repeated game is $\overline{v}(s) = R$. Indeed, if $x \in X^{**}$ then $\overline{u}(s, x, \sigma) \geq R$ for any $\sigma \in \Sigma$. Hence, $\overline{v}(s) \geq R$. On the other hand, if $b \in B(s)$ is such that the minimum in (8) is attained at b, then $\overline{u}(s, \pi, b^\infty) \leq R$ for any $\pi \in \Pi$. Here, b^∞ is the stationary strategy for player 2 that always chooses action b. Hence, $\overline{v}(s) \leq R$.

It follows from the arguments above that $\gamma^*(s) = R$ and that each maximally mixed strategy $x \in X^{**}$ is effective, and optimal for the limsup payoff. ◇

Example 5 (*When Player 1 has an Optimal Strategy*)**.**

	L	R
T	2	0^*
B	0^*	1^*

We examine this game with the limsup payoff. The notation is similar to those of the previous examples.

Let s denote the non-absorbing state. Player 2 can guarantee that the limsup payoff is at most ε, for any $\varepsilon \in (0, 1)$, by playing the stationary strategy that chooses action L with probability $1 - \varepsilon$ and action R with probability ε. Hence, $\overline{v}(s) = 0$. Note that any strategy of player 1 is optimal.

The corresponding matrix game $M(s)$ is

	L	R
T	0	0
B	0	1

Hence, $X^*(s) = X(s)$, $X^{**}(s) = \{(p, 1-p) | p \in (0, 1)\}$ and $Y^*(s) = \{(1, 0)\}$. Consequently, due to payoff 2 in the cell (T, L) in the original game, we obtain $\gamma^*(s) = 2$. It is easy to check that any maximally mixed strategy $x \in X^{**}$ is effective. Indeed, $y = (1, 0)$ is a (unique) best response for player 2, and the ergodic sets for (x, y) are the singletons consisting of the absorbing states. $\diamond$

Example 6 (*When Player 1 has no Optimal Strategy*)**.** We revisit the game of Example 1 with the limsup payoff.

	L	R
T	1	0
B	0^*	1^*

As before, s denotes the non-absorbing state. Recall that $\overline{v}(s) = 1$.

The corresponding matrix game $M(s)$ is

	L	R
T	1	1
B	0	1

Hence, $X^*(s) = \{(1, 0)\}$, $X^{**}(s) = \{(1, 0)\}$ and $Y^*(s) = Y(s)$. Consequently, due to payoff 0 in the cell (T, R) in the original game, we obtain $\gamma^*(s) = 0$. The unique maximally mixed strategy in X^{**} is effective. But, it is not ε-optimal for $\varepsilon \in [0, 1)$. $\diamond$

6. The proof of Theorem 2

We start with a general remark regarding the proof of Theorem 2. Consider a zero-sum stochastic game G with the liminf payoff. By making use of lemma 4.9 in [23], we could conclude that there exists a stationary strategy x^* for player 1 such that, for all strategies σ for player 2 and all initial states s, we have $u(s, x^*, \sigma) \geq \mathbb{E}_{(s,x^*,\sigma)}[\lim_t \underline{v}(s_t)]$. Suppose that, in addition, x^* can be chosen so that, for every state s, the mixed action $x^*(s)$ is optimal in the matrix game $M(s)$. Then, for all strategies σ for player 2 and all initial states s, the stochastic process $\underline{v}(s_t)$ of successive values is a bounded submartingale under $\mathbb{P}_{(s,x^*,\sigma)}$. Hence, $\mathbb{E}_{(s,x^*,\sigma)}[\lim_t \underline{v}(s_t)] \geq \underline{v}(s)$. Therefore such a strategy x^* would be optimal for player 1 in the game G.

In an attempt to find such a strategy, provided that player 1 has a subgame-optimal strategy, it is natural to restrict player 1 to mixed actions at each state that are optimal in the corresponding matrix game and then to apply Secchi's result. However, there is a technical difficulty with such an approach. As is explained below, the restricted game is not a well-defined stochastic game in the usual sense. Much of the proof of Theorem 2 is devoted to overcoming this technical difficulty.

We define two conditions on a zero-sum stochastic game that will play an important role in the proof.

Condition C1. The reward function only depends on the state, and thus not on the actions chosen by the players: $r(s, a, b) = r(s, a', b')$ for all states $s \in S$, and actions $a, a' \in A(s)$ and $b, b' \in B(s)$.

Condition C2. In every state, different actions of player 1 lead to different states: for all states $s \in S$ and distinct actions $a, a' \in A(s)$, the sets

$$\{w \in S \mid \exists b \in B(s) : \ p(w|s, a, b) > 0\} \quad \text{and} \quad \{w \in S \mid \exists b \in B(s) : \ p(w|s, a', b) > 0\}$$

are disjoint.

Consider an arbitrary zero-sum stochastic game G with the liminf payoff. Let $X^*(s)$ for every $s \in S$, and Π^* be as in Section 5.

The related stochastic game G^*: We define a closely related stochastic game G^*. The idea behind G^* is to restrict player 1 to only using mixed actions in $X^*(s)$, that is, optimal mixed actions in the matrix game $M(s)$, in state s. The subtlety of this approach lies in the fact that restricting player 1's set of mixed strategies to $X^*(s)$ does not lead to a well-defined stochastic

game, since player 2 does not observe mixtures played by player 1, only the realization of these mixtures. To circumvent this difficulty we introduce an auxiliary game G^* where the extreme points of $X^*(s)$ are declared player 1's pure actions, and we impose the assumption that player 2 observes them. All this is done so that we are able to apply the result of Secchi [23] on liminf stochastic games.

For every state $s \in S$, let $E(s)$ denote the set of extreme points of $X^*(s)$. Since $X^*(s)$ is a nonempty polytope, the set $E(s)$ is nonempty and finite. The state space of the new stochastic game G^* is S. In state $s \in S$, the action space is $E(s)$ for player 1 and $B(s)$ for player 2. For states $s \in S$ and actions $e \in E(s)$, $b \in B(s)$, the reward is given by

$$r^*(s, e, b) = \sum_{a \in A(s)} e(a) \cdot r(s, a, b),$$

and the transition is given by

$$p^*(w|s, e, b) = \sum_{a \in A(s)} e(a) \cdot p(w|s, a, b) \quad \forall w \in S.$$

We remark that if Condition C1 is satisfied, then $r^*(s, e, b)$ does not depend on e and b and coincides with $r(s, a, b)$ for each $a \in A(s)$ and $b \in B(s)$.

During the play of the game G^*, as usual, the players observe the current state and the actions chosen in the current state. That is, if the current state is s, the players observe s, and after choosing actions, they observe the action chosen in $E(s)$ by player 1 and the action chosen in $B(s)$ by player 2. In this game G^*, the set $A(s)$ has no special meaning, and is merely used to define the rewards and the transitions. We denote by $\underline{v}^*(s)$ the liminf value of G^*. The following lemma provides a connection between the liminf value of G^* and the original game G.

Lemma 7. *Consider a zero-sum stochastic game G with the liminf payoff that satisfies Conditions C1 and C2. Then, for every state $s \in S$ we have*

$$\underline{v}^*(s) = \sup_{\pi \in \Pi^*} \inf_{\sigma \in \Sigma} \underline{u}(s, \pi, \sigma) = \inf_{\sigma \in \Sigma} \sup_{\pi \in \Pi^*} \underline{u}(s, \pi, \sigma).$$

Proof. Consider a zero-sum stochastic game G with the liminf payoff that satisfies Conditions C1 and C2. Let W denote the set of all sequences of the form $(s_0, b_0, \ldots, s_{t-1}, b_{t-1}, s_t)$, where $s_k \in S$ for all $k \in \{0, \ldots, t\}$, and $b_k \in B(s_k)$ for all $k \in \{0, \ldots, t-1\}$.

Further notation and definitions for the game G: By Condition C2, for all states $s, s' \in S$ there is at most one action $a \in A(s)$ such that $p(s'|s, a, b) > 0$ for some $b \in B(s)$. Whenever such an action exists, we denote it by $a_{s,s'}$. Thus, if the play moves from state s to state s', then the players can conclude that player 1 played action $a_{s,s'}$ in state s.

Let $\xi : H \to W$ denote the mapping that, to each history $h \in H$, assigns the sequence $\xi(h) \in W$ that arises by erasing the actions of player 1. Note that by Condition C2, ξ is a bijection.[6] This allows us to consider W as the set of histories in the game G, and as the domain of strategies in Π^* (or even Π) and Σ.

Further notation and definitions for the game G^*: We denote by $\underline{u}^*$ the liminf payoff function for G^*. Let α and β denote strategies in G^* for players 1 and 2 respectively, and let Φ and Ψ denote the sets of strategies in G^* for players 1 and 2 respectively.

Let H^* denote the set of histories in the game G^*. A history in H^* is thus of the form

$$h = (s_0, e_0, b_0, \ldots, s_{t-1}, e_{t-1}, b_{t-1}, s_t),$$

where $s_k \in S$ for all $k \in \{0, \ldots, t\}$, and $e_k \in E(s_k)$ and $b_k \in B(s_k)$ for all $k \in \{0, \ldots, t-1\}$.

Let $\xi^* : H^* \to W$ denote the mapping that, to each $h \in H^*$, assigns the sequence $\xi^*(h) \in W$ that arises by erasing the actions of player 1.

For player 1, let Φ^* denote the set of strategies in Φ that prescribe a mixed action only depending on the sequence of states and the sequence of actions played by player 2. Thus, these strategies do not make use of the actions played by player 1. Consequently, the domain of strategies in Φ^* can be identified with W.

For player 2, similarly, let Ψ^* denote the set of strategies in Ψ that prescribe a mixed action only depending on the sequence of states and the sequence of actions played by player 2. Consequently, the domain of strategies in Ψ^* can be identified with W. Due to this, we can also identify Ψ^* with Σ.

Identification of strategies between G **and** G^*: As discussed above, the domain of the strategy sets Φ^* and Π^* for player 1 can be identified with W. We can therefore define a mapping $\rho_1 : \Phi^* \to \Pi^*$ such that, for all $\alpha \in \Phi^*$, the strategy $\rho_1(\alpha)$ is the unique strategy $\pi \in \Pi^*$ with the following property: for each sequence $h = (s_0, b_0, \ldots, s_{t-1}, b_{t-1}, s_t)$ in W we have

$$\pi(h)(a) = \sum_{e \in E(s_t)} \alpha(h)(e) \cdot e(a) \qquad \forall a \in A(s_t).$$

[6] Up to histories in H and sequences in W that are not consistent with the transitions of the game.

Intuitively, for any sequence h in W, the probability of any action $a \in A(s_t)$ under $\pi(h)$ is equal to the probability of a under $\alpha(h)$, i.e., that action a is chosen by $e \in E(s_t)$, after e being drawn according to $\alpha(h)$. The mapping ρ_1 is surjective, but generally not one-to-one.

For completeness, for player 2 we denote the identity mapping $\Psi^* \to \Sigma$ by ρ_2.

We then have for all states $s \in S$, and all strategies $\alpha \in \Phi^*$ and $\beta \in \Psi^*$ that

$$\underline{u}^*(s, \alpha, \beta) = \underline{u}(s, \rho_1(\alpha), \rho_2(\beta)). \tag{9}$$

This equality relies on the fact that (α, β) and $(\rho_1(\alpha), \rho_2(\beta))$ generate the same probability measure on W, thus using C1 and the fact that for a fixed state s, the payoff $r^*(s, e, b)$ is independent of e and b, they generate the same probability measure on $r(Z)^\infty$.

Hence, by surjectivity of ρ_1 and ρ_2, we have for all states $s \in S$

$$\inf_{\beta\in\Psi^*} \sup_{\alpha\in\Phi^*} \underline{u}^*(s, \alpha, \beta) = \inf_{\beta\in\Psi^*} \sup_{\alpha\in\Phi^*} \underline{u}(s, \rho_1(\alpha), \rho_2(\beta)) = \inf_{\sigma\in\Sigma} \sup_{\pi\in\Pi^*} \underline{u}(s, \pi, \sigma). \tag{10}$$

The main body of the proof. Fix an initial state $s \in S$.

Part A. First, we obviously have

$$\sup_{\pi\in\Pi^*} \inf_{\sigma\in\Sigma} \underline{u}(s, \pi, \sigma) \le \inf_{\sigma\in\Sigma} \sup_{\pi\in\Pi^*} \underline{u}(s, \pi, \sigma). \tag{11}$$

(For this inequality to hold, we only need that $\underline{u}$ is a real-valued function.)

Part B. Now we prove

$$\underline{v}^*(s) \le \sup_{\pi\in\Pi^*} \inf_{\sigma\in\Sigma} \underline{u}(s, \pi, \sigma). \tag{12}$$

Let $\varepsilon > 0$. By theorem 4.16 in [23], player 1 has a stationary ε-optimal strategy α in the game G^*. Denote $x = \rho_1(\alpha)$. Thus, $x \in X^*$. Let $y \in Y$ be a stationary best response of player 2 to x (such a strategy y is a best response both in G and in G^*). Then, by using (9),

$$\begin{aligned} \underline{v}^*(s) - \varepsilon &\le \underline{u}^*(s, \alpha, y) \\ &= \underline{u}(s, x, y) \\ &= \inf_{\sigma\in\Sigma} \underline{u}(s, x, \sigma) \\ &\le \sup_{\pi\in\Pi^*} \inf_{\sigma\in\Sigma} \underline{u}(s, \pi, \sigma). \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, we have shown (12).

Part C. Now we argue that

$$\underline{v}^*(s) \ge \inf_{\sigma\in\Sigma} \sup_{\pi\in\Pi^*} \underline{u}(s, \pi, \sigma). \tag{13}$$

As the game G satisfies Condition C1 by assumption, so does the game G^*. As remarked at the end of Section 3.1, under Condition C1, player 2 has an ε-optimal strategy, for every $\varepsilon > 0$, which does not take the actions chosen by player 1 into account. That is, player 2 has an ε-optimal strategy in G^*, for every $\varepsilon > 0$, within the set Ψ^*. Hence,

$$\begin{aligned} \underline{v}^*(s) &= \inf_{\beta\in\Psi} \sup_{\alpha\in\Phi} \underline{u}^*(s, \alpha, \beta) \\ &= \inf_{\beta\in\Psi^*} \sup_{\alpha\in\Phi} \underline{u}^*(s, \alpha, \beta) \\ &\ge \inf_{\beta\in\Psi^*} \sup_{\alpha\in\Phi^*} \underline{u}^*(s, \alpha, \beta) \\ &= \inf_{\sigma\in\Sigma} \sup_{\pi\in\Pi^*} \underline{u}(s, \pi, \sigma), \end{aligned}$$

where the last equality is (10), and this proves (13).

By (11), (12) and (13), the proof of Lemma 7 is complete. □

Lemma 8. *Consider a zero-sum stochastic game G with the liminf payoff that satisfies Conditions C1 and C2. If player 1 has a subgame-optimal strategy, then he has a stationary optimal strategy as well.*

Proof. Assume that the zero-sum stochastic game G with the liminf payoff satisfies Conditions C1 and C2. Let π^* be a subgame-optimal strategy for player 1 in the game G.

Due to Condition C1, the reward function r of G and hence also the reward function r^* of G^* only depend on the state. For the sake of exposition, let $r(s) = r^*(s)$ denote the reward for every state $s \in S$.

Step 1: The strategy π^* is a member of the restricted set of strategies Π^*.

Proof of step 1: An argument by contradiction, similar to that for step 1 in the proof of Theorem 5, shows that an optimal strategy in the stochastic game at any state s must begin with an optimal mixed action in the matrix game $M(s)$. Because π^* is optimal in every subgame, the action $\pi^*(h)$ must be optimal in $M(s_h)$ for every history h. Thus $\pi^*(h) \in X^*(s_h)$ for all h, and therefore $\pi^* \in \Pi^*$.

Step 2: We have for every state $s \in S$

$$\sup_{\pi \in \Pi^*} \inf_{\sigma \in \Sigma} \underline{u}(s, \pi, \sigma) \geq \underline{v}(s).$$

Proof of step 2: This is by step 1 and the optimality of π^*.

Step 3: For every $\pi \in \Pi^*, \sigma \in \Sigma$, and $s \in S$, the stochastic process $\{\underline{v}(s_t)\}$ converges $\mathbb{P}_{(s,\pi,\sigma)}$-almost surely and $\mathbb{E}_{(s,\pi,\sigma)}[\lim_{t\to\infty} \underline{v}(s_t)] \geq \underline{v}(s)$.

Proof of step 3: The process $\{\underline{v}(s_t)\}$ is a $\mathbb{P}_{(s,\pi,\sigma)}$-submartingale because π selects mixed actions optimal in $M(s_h)$ at every h. The process is bounded because the reward function r is bounded. So the almost sure convergence of $\{\underline{v}(s_t)\}$ follows from a martingale convergence theorem. Also, by the dominated convergence theorem and the fact that submartingales are nondecreasing in expectation:

$$\mathbb{E}_{(s,\pi,\sigma)}[\lim_{t\to\infty} \underline{v}(s_t)] = \lim_{t\to\infty} \mathbb{E}_{(s,\pi,\sigma)}[\underline{v}(s_t)] \geq \underline{v}(s).$$

Step 4: Recall that $\underline{v}^*(s)$ denotes the liminf value of the game G^*. We claim for the original game G that there exists a stationary strategy $x^* \in X^*$ for player 1 such that, for all stationary strategies $y \in Y$ and all $s \in S$,

$$\mathbb{E}_{(s,x^*,y)}[\liminf_{t\to\infty} r(s_t)] \geq \mathbb{E}_{(s,x^*,y)}[\liminf_{t\to\infty} \underline{v}^*(s_t)].$$

Proof of step 4: By applying lemma 4.9 of Secchi [23] to the game G^*, there exists a stationary strategy $\widetilde{e}$ for player 1 in G^*, i.e. $\widetilde{e} \in \times_{s\in S} \Delta(E(s))$ where $\Delta(E(s))$ stands for the set of probability measures on $E(s)$, such that for all stationary strategies $y \in Y$ and all $s \in S$

$$\mathbb{P}_{(s,\widetilde{e},y)}[r(s_t) \geq \underline{v}^*(s_t) \text{ for all but finitely many } t] = 1.$$

Let $x^* \in X^*$ be the stationary strategy for player 1 in the game G defined by letting for each state $s \in S$ and each action $a \in A(s)$

$$x^*(s, a) = \sum_{e \in E(s)} \widetilde{e}(s, e) \cdot e(a).$$

Thus, the probability that x^* places on action a in state s is exactly the probability that one obtains if $e \in E(s)$ is drawn from $\widetilde{e}$ in state s, and subsequently a is drawn from e.[7] Since stationary strategies only take the current state into account when choosing an action, we obtain for all stationary strategies $y \in Y$ and all $s \in S$

$$\mathbb{P}_{(s,x^*,y)}[r(s_t) \geq \underline{v}^*(s_t) \text{ for all but finitely many } t] = 1.$$

Hence, for all stationary strategies $y \in Y$ and all $s \in S$

$$\mathbb{P}_{(s,x^*,y)}[\liminf_{t\to\infty} r(s_t) \geq \liminf_{t\to\infty} \underline{v}^*(s_t)] = 1,$$

which establishes the claim of step 4.

Conclusion of the proof of Lemma 8: Let x^* be a stationary strategy as in step 4, and let $y \in Y$ be a best response for player 2 against x^*. Then, by step 4, we have for all $s \in S$ that

$$\underline{u}(s, x^*, y) = \mathbb{E}_{(s,x^*,y)}[\liminf_{t\to\infty} r(s_t)] \geq \mathbb{E}_{(s,x^*,y)}[\liminf_{t\to\infty} \underline{v}^*(s_t)].$$

In view of Lemma 7 and step 2, we have for every state $w \in S$ that

$$\underline{v}^*(w) = \sup_{\pi \in \Pi^*} \inf_{\sigma \in \Sigma} \underline{u}(w, \pi, \sigma) \geq \underline{v}(w).$$

Hence, for all $s \in S$

$$\underline{u}(s, x^*, y) \geq \mathbb{E}_{(s,x^*,y)}[\liminf_{t\to\infty} \underline{v}(s_t)]. \tag{14}$$

[7] By using the mapping ρ_1 from the proof of Lemma 7, we have $x^* = \rho_1(\widetilde{e})$.

Then, step 3 implies for all $s \in S$ that

$$\underline{u}(s, x^*, y) \geq \underline{v}(s).$$

Since y is a best response to x^*, we conclude that x^* is a stationary optimal strategy for player 1 in the game G. □

In the proof above, the conditions on π in step 3 and x^* in (14) are two-person versions of the conditions [10] called "thrifty" and "equalizing" in their study of one-person gambling problems with a limsup payoff. The corresponding conditions for gambling problems with a liminf payoff are in [25].

Proof of Theorem 2. Consider a zero-sum stochastic game G with the liminf payoff.

Now we construct another zero-sum stochastic game G^E, which can be seen as the game G on an extended state space. The state space of G^E is

$$S^E = \{(s, a, b, w) \mid s, w \in S,\ a \in A(s),\ b \in B(s)\}.$$

The idea is that state (s, a, b, w) describes the situation in the original game G when, at the previous period, the play was in state s and the players chose actions a and b, and then w became the current state. Thus, the game G^E is just like the game G with the modification that the players can conclude from the current state what the previous state was and which actions the players chose there.

We describe the game G^E more formally. In state $s^E = (s, a, b, w) \in S^E$, the action spaces are $A^E(s^E) = A(w)$ for player 1 and $B^E(s^E) = B(w)$ for player 2, and the reward is equal to $r^E(s^E, a', b') = r(s, a, b)$ for all actions $a' \in A^E(s^E)$ and $b' \in B^E(s^E)$. Intuitively, in the game G^E the reward is equal to the reward in G at the previous period. In state $s^E = (s, a, b, w) \in S^E$, the transition for actions $a' \in A^E(s^E)$ and $b' \in B^E(s^E)$ is as follows: for states of the form $w^E = (w, a', b', w')$, where $w' \in S$,

$$p^E(w^E | s^E, a', b') = p(w' | w, a', b'),$$

and the transition probabilities are zero for all other states in S^E.

Note that by construction, the game G^E satisfies Conditions C1 and C2.

Assume that the game G starts in initial state $s \in S$. Choose an initial state (w, a, b, s) for G^E, where $w \in S$, $a \in A(w)$ and $b \in B(w)$ are arbitrary. Given these initial states, there is a bijection between the histories in G and G^E, where a history

$$h = (s_0, a_0, b_0, s_1, a_1, b_1, \ldots, s_{t-1}, a_{t-1}, b_{t-1}, s_t)$$

in the game G, where $s_0 = s$, corresponds to the history

$$h^E = ((w, a, b, s_0), a_0, b_0, (s_0, a_0, b_0, s_1), a_1, b_1, \ldots, (s_{t-2}, a_{t-2}, b_{t-2}, s_{t-1}), a_{t-1}, b_{t-1}, (s_{t-1}, a_{t-1}, b_{t-1}, s_t))$$

in the game G^E. Let $r_k = r(s_k, a_k, b_k)$ for $k \in \{0, \ldots, t-1\}$. Then, h induces rewards $(r_0, \ldots, r_{t-1})$, whereas h^E induces rewards $(r(w, a, b), r_0, \ldots, r_{t-2})$. That is, in the game G^E, there is a one period delay in the rewards.

Given the bijection between histories, there is also a bijection between strategies in G and G^E. Furthermore, for any strategies π and σ in G and corresponding strategies π^E and σ^E in G^E, we have $\underline{u}(s, \pi, \sigma) = \underline{u}^E((w, a, b, s), \pi^E, \sigma^E)$. Consequently, the liminf values of the games G and G^E satisfy

$$\underline{v}^E(w, a, b, s) = \underline{v}(s).$$

By assumption, player 1 has a subgame-optimal strategy π in G. Due to construction, player 1 has a subgame-optimal strategy π^E in G^E. As the game G^E satisfies Conditions C1 and C2, we can conclude by Lemma 8 that player 1 has a (subgame-optimal) stationary optimal strategy x^E in G^E. This strategy in turn induces in the game G a subgame-optimal strategy for player 1 under which the prescribed mixed actions only depend on the current state and on the state and the actions chosen at the previous period. This completes the proof of Theorem 2. □

Example 7. To illustrate the construction of the game G^E, let G be the Big Match (cf. Example 1). By having the payoff and the transition in each cell, the game G can be represented as

	L	R
T	$1, s$	$0, s$
B	$0, \ell$	$1, w$

state s

	b
a	$0, \ell$

state ℓ

	b
a	$1, w$

state w

The extended game G^E is then

	L	R
T	$1, (s, T, L, s)$	$1, (s, T, R, s)$
B	$1, (s, B, L, \ell)$	$1, (s, B, R, w)$

state (s, T, L, s)

	L	R
T	$0, (s, T, L, s)$	$0, (s, T, R, s)$
B	$0, (s, B, L, \ell)$	$0, (s, B, R, w)$

state (s, T, R, s)

	b
a	$0, (\ell, a, b, \ell)$

state (s, B, L, ℓ)

	b
a	$0, (\ell, a, b, \ell)$

state (ℓ, a, b, ℓ)

$$a \;\overset{b}{\boxed{\begin{array}{c}1,(w,a,b,w)\\ \text{state } (s,B,R,w)\end{array}}} \qquad a \;\overset{b}{\boxed{\begin{array}{c}1,(w,a,b,w)\\ \text{state } (w,a,b,w)\end{array}}}$$

Let s be the initial state in the game G, and let (s,T,L,s) be the initial state in the game G^E (the arguments below are similar if (s,T,R,s) is the initial state in the game G^E). Suppose that the history in G is

$$h = (s,T,R,s,T,R,s,T,L,s,T,R,s,B,R,w,a,b,w,a,b,w).$$

Then, the sequence of rewards is $r(h) = (0,0,1,0,1,1,1)$. The corresponding history in the game G^E is

$$h^E = ((s,T,L,s),T,R,(s,T,R,s),T,R,(s,T,R,s),T,L,(s,T,L,s),T,R,\\ (s,T,R,s),B,R,(s,B,R,w),a,b,(w,a,b),a,b,(w,a,b)),$$

with sequence of rewards $r^E(h^E) = (1,0,0,1,0,1,1)$. This sequence $r^E(h^E)$ starts with reward 1 due to the choice of (s,T,L,s) as the initial state ((s,T,R,s) would induce reward 0), and then coordinate $t+1$ of $r^E(h^E)$ is equal to coordinate t of $r(h)$. So, in the game G^E, as we mentioned earlier, the rewards are received one period later than in G.

7. Extensions

In this section, we discuss extensions of the main results.

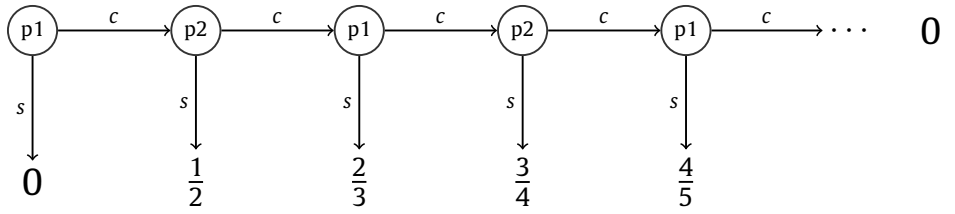
7.1. Games with countably infinite state spaces

The following example demonstrates that, if the state space is countably infinite, then the existence of an optimal strategy does not imply the existence of a subgame-optimal strategy, neither for the limsup nor for the liminf payoff. In particular, Theorem 1 cannot be extended to games with countably infinite state spaces. It remains an open question whether Theorem 2 is valid in games with a countably infinite state space.

Example 8. Consider the following game. The state space is $\mathbb{N} \times \{c,s\}$. The states can be described as follows:

- In each state (n,c), where n is even, player 1 has two actions c and s and player 2 has only one action. If player 1 chooses action c, then the payoff is 0 and the play moves to state $(n+1,c)$. If player 1 chooses action s, then the payoff is 0 and the play moves to state (n,s).
- In each state (n,c), where n is odd, player 2 has two actions c and s and player 1 has only one action. If player 2 chooses action c, then the payoff is 0 and the play moves to state $(n+1,c)$. If player 2 chooses action s, then the payoff is 0 and the play moves to state (n,s).
- Each state (n,s) is absorbing, and the payoff is $\frac{n}{n+1}$.

Notice that this game has perfect information, i.e., in every state only one player has more than one action (we can assume that each player has only one action in the absorbing states). Moreover, the limsup and the liminf payoffs are equal to each other, for any pair of strategies. In fact, this game with either payoff is equivalent to the following centipede game:

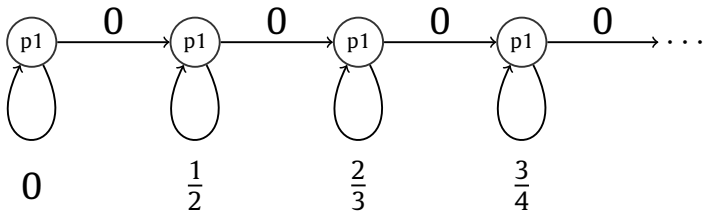


Take an initial state (n,c), where n is even. Player 1 is the active player at this state. One can easily verify for this initial state that (1) the value is $\frac{n+1}{n+2}$, (2) it is optimal for player 1 to play action c in state (n,c) and action s in state $(n+2,c)$, (3) it is optimal for player 2 to play action s in state $(n+1,c)$. Note that any optimal strategy of player 1 requires to play action c in state (n,c).

Consequently, player 1 has an optimal strategy in the game. However, player 1 has no subgame-optimal strategy, as playing action c in every state (n,c), where n is even, only gives payoff 0 against the strategy of player 2 that always chooses action c. ◇

The next example is motivated by a game in [10, example 3 in section 3.9].

Example 9. Consider the following game with the limsup payoff. The state space is $\mathbb{N}$. Player 2 is a dummy, i.e., he has only one action in each state. In state $n \in \mathbb{N}$, player 1 has two actions. One action gives reward 0 and leads to state $n+1$, whereas the other action gives reward $\frac{n}{n+1}$ and keeps the play in state n. The game can be represented as follows:



In this game, player 1 has a subgame-optimal strategy. Indeed, staying in every state for exactly 2 periods constitutes a pure subgame-optimal strategy, and choosing each action with probability 0.5 in each state constitutes a stationary subgame-optimal strategy. However, player 1 does not possess a pure stationary optimal strategy, which might be surprising given the fact that this is essentially a one-player game (Markov Decision Problem).

7.2. Games with countably infinite action spaces

In this section we consider games with a finite state space but possibly infinite action spaces in which player 2 is dummy, i.e. he has only one action in each state. As mentioned above, these games can be seen as one-player games.

For the limsup payoff, Dubins and Savage [10, theorem 3, p. 59] showed that, if the player has an optimal strategy, then he also has a stationary optimal strategy.

The example below shows that this is not true for the liminf payoff even if there is a subgame-optimal strategy. It is true however that in the example there is an optimal strategy for which the action chosen at each period depends only on the current state together with the state and action of the previous period. Thus there is an optimal strategy of the type introduced in Theorem 2.

We remark that if both the state and action spaces are finite, there is always a stationary optimal strategy in both the limsup and liminf cases, cf. [10, theorem 1, p. 58] and [25].

Example 10. Consider the following one-player game with the liminf payoff. The state space is $S = \{0, 1\}$. In state 0, there is only one action, which gives reward 0 and induces transition to state 1 with probability 1. In state 1, the action space is $\{a_n \mid n = 2, 3, \ldots\}$, where the action a_n, for each n, gives reward 1 and induces transition probabilities $p(0|1, a_n) = \frac{1}{n}$ and $p(1|1, a_n) = 1 - \frac{1}{n}$. Let W denote the set of all sequences in S that are eventually equal to 1, i.e. in which 0 only appears finitely many times. Thus, the player receives a liminf payoff of 1 if the sequence of states visited during the course of the play is in W, and receives a liminf payoff of 0 otherwise.

For simplicity, assume that state 1 is the initial state. To construct an optimal strategy, first choose a sequence of integers $n_1, n_2, \ldots$, each at least 2, such that $q = \Pi_{k=1}^{\infty}(1 - \frac{1}{n_k})$ is strictly positive. Then let π be the strategy that uses $a_{n_1}, a_{n_2}, \ldots$ in order until and if ever state 0 is reached, and upon each return to state 1 starts over again using $a_{n_1}, a_{n_2}, \ldots$ as before. With respect to π, the probability of staying at state 1 forever is q, and the probability of inducing a sequence of states in W is

$$q + (1-q) \cdot q + (1-q)^2 \cdot q + \cdots = 1.$$

It follows that the liminf value is equal to 1 and that π is optimal. It is not difficult to see that π induces payoff 1 in every subgame and thus π is even subgame-optimal.

Further notice that no stationary strategy can be optimal, whether pure or randomized. Indeed, the probability under any stationary strategy of staying at state 1 forever is zero, and it follows that the probability of inducing a sequence of states in W is also 0.

Here is another description of the optimal strategy defined above:

- If the current state is 1 and the previous state was 0, play the action a_{n_1}.
- If the current state is 1, the previous state was 1 and the previous action was a_{n_k}, play the action $a_{n_{k+1}}$.
- If the current state is 0, play the only action available.

Thus, the example shows that there may exist an optimal strategy as in Theorem 2, even though there does not exist a stationary optimal strategy. ◇

7.3. Limsup and liminf of the expected rewards

Under our definition of the evaluation $\overline{u}(s, \pi, \sigma)$ of the pair of strategies (π, σ) in the initial state s, one computes the expectation of the limit superior of the sequence of rewards. In this section we briefly examine another evaluation criterion thereby one takes the limit superior of the expectation of the rewards.

For an initial state s and pair of strategies (π, σ), we define

$$\overline{\overline{u}}(s, \pi, \sigma) = \limsup_{t\to\infty} \mathbb{E}_{(s,\pi,\sigma)}(r_t),$$

where $\mathbb{E}_{(s,\pi,\sigma)}(r_t)$ denotes the expectation of the reward at period t under (s, π, σ). Fatou's lemma implies that $\overline{\overline{u}}(s, \pi, \sigma) \leq \overline{u}(s, \pi, \sigma)$.[8] Similarly, we define

$$\underline{\underline{u}}(s, \pi, \sigma) = \liminf_{t\to\infty} \mathbb{E}_{(s,\pi,\sigma)}(r_t).$$

The following example shows that Theorem 1 does not extend to the case of the payoff $\overline{\overline{u}}$. It remains an open question whether Theorem 2 could be extended to $\underline{\underline{u}}$.

Example 11. Consider the following game, with the notation similar to those of the previous examples:

	L	R
T	1	0
B	0	1^*

With respect to the payoff $\overline{\overline{u}}$, player 1 has a Markov optimal strategy, but no stationary optimal strategy. Thus, Theorem 1 does not extend to the payoff $\overline{\overline{u}}$.

As before, let s denote the non-absorbing state. Let π be the Markov strategy which prescribes to play action T with probability $\frac{t}{t+1}$ and action B with probability $\frac{1}{t+1}$ when in state s in period t. The idea behind this strategy is that if player 2 tries to play action L infinitely often, then $\overline{\overline{u}}$ gives payoff 1, whereas if player 2 tries to play action R at each period $t \geq T$, for some $T \in \mathbb{N}$, then play will absorb in entry (B, R) with probability 1, and $\overline{\overline{u}}$ gives payoff 1 again. Thus, one can show[9] that π guarantees a payoff of 1 under $\overline{\overline{u}}$ and is thus optimal. However, it is easy to see that player 1 has no stationary optimal strategy for $\overline{\overline{u}}$. ◇

References

[1] R. Amir, Stochastic games and economics and related fields: an overview, in: A. Neyman, S. Sorin (Eds.), Stochastic Games and Applications, in: NATO Science series C, vol. 570, Kluwer Academic, Dordrecht, 2003.
[2] K.R. Apt, E. Grädel (Eds.), Lectures in Game Theory for Computer Scientists, Cambridge University Press, 2011.
[3] D. Blackwell, On stationary policies, J. Roy. Statist. Soc. 133 (1971) 33–37.
[4] D. Blackwell, T.S. Ferguson, The big match, Ann. Math. Statist. 33 (1968) 159–163.
[5] V. Bruyère, Computer aided synthesis: a game-theoretic approach, Lecture Notes in Comput. Sci. 10396 (2017) 3–35.
[6] K. Chatterjee, L. Doyen, T.A. Henzinger, A Survey of stochastic games with limsup and liminf objectives, in: Proceedings of ICALP(2), 2009.
[7] K. Chatterjee, T.A. Henziger, A survey of stochastic ω-regular games, J. Comput. System Sci. 78 (2012) 394–413.
[8] K. Chatterjee, M. Jurdzinski, T.A. Henzinger, Quantitative stochastic parity games. Proceedings of the Symposium on Discrete Algorithms (SODA), 2004.
[9] L. de Alfaro, Formal Verification of Probabilistic Systems (Ph.D. thesis), Stanford University, 1997.
[10] L.E. Dubins, L.J. Savage, How To Gamble if You Must: Inequalities for Stochastic Processes, Dover editions in 1976 and 2014, McGraw-Hill, New York, 1965.
[11] J. Flesch, F. Thuijsman, K. Vrieze, Simplifying optimal strategies in stochastic games, SIAM J. Control Optim. 36 (1998) 1331–1347.
[12] J. Flesch, F. Thuijsman, K. Vrieze, Markov strategies are better than stationary strategies, Int. Game Theory Rev. 1 (1999) 9–31.
[13] D. Gillette, Stochastic games with zero stop probabilities, in: M. Dresher, A.W. Tucker, P. Wolfe (Eds.), Contributions to the Theory of Games, Vol. III, in: Annals of Mathematics Studies, 39, Princeton University Press, Princeton, NJ, 1957, pp. 179–187.
[14] H. Gimbert, Jeux Positionnels (Ph.D. thesis), Université Paris 7, 2006.
[15] H. Gimbert, Pure stationary optimal strategies in Markov decision processes, Lecture Notes in Comput. Sci. 4393 (2007) 200–211.
[16] H. Gimbert, J. Renault, S. Sorin, X. Venel, A. Zielonka, On values of repeated games with signals, Ann. Appl. Probab. 26 (2016) 402–424.
[17] A. Maitra, W. Sudderth, An operator solution of stochastic games, Israel J. Math. 78 (1992) 33–49.
[18] A. Maitra, W. Sudderth, Borel stochastic games with the limsup payoff, Ann. Probab. 21 (1993) 861–885.
[19] A. Maitra, W. Sudderth, Discrete Gambling and Stochastic Games, Springer, 1996.
[20] A. Maitra, W. Sudderth, Stochastic games with Borel payoffs, in: A. Neyman, S. Sorin (Eds.), Stochastic Games and Applications, Springer, Dordrecht, The Netherlands, 2003, pp. 367–373.
[21] D.A. Martin, The Determinacy of Blackwell Games, J. Symbolic Comput. 63 (1998) 1565–1581.
[22] M. Orkin, On stationary policies–the general case, Ann. Statist. 2 (1974) 219–222.
[23] P. Secchi, On the existence of good stationary strategies for nonleavable stochastic games, Internat. J. Game Theory 27 (1998) 61–81.
[24] L.S. Shapley, Stochastic games, Proceedings of the National Acadamy of Sciences of the United States of America. 39 (1953) 1095–1100.
[25] W. Sudderth, Gambling problems with a limit inferior payoff, Math. Oper. Res. 8 (1983) 287–297.
[26] W. Sudderth, Optimal Markov Strategies. Preprint, 2016.

[8] Note that the inequality can be strict, as the following game illustrates. There are only two states, and one action for each player in each state. The reward is -1 in one state, and 1 in the other state. The transitions bring the game to both states with probability 0.5. Then, for either initial state, $\overline{\overline{u}}$ gives payoff 0, whereas $\overline{u}$ gives payoff 1.

[9] One way of proving this is to use [11,12]. There it is shown that π guarantees the payoff of 1 with respect to the expected average reward evaluation

$$u'(s, \pi, \sigma) = \limsup_{T\to\infty} \frac{1}{T+1} \mathbb{E}_{(s,\pi,\sigma)}(r_0 + \cdots + r_T).$$

Since $\overline{\overline{u}} \geq u'$, π guarantees 1 for $\overline{\overline{u}}$ as well. We remark that Flesch et al. consider this game with payoff 2 in entry (B, R) for illustrative purposes, but this makes no difference in our case.